PREDICTION OF PNEUMONIA USING BIG DATA, DEEP LEARNING AND MACHINE LEARNING TECHNIQUES

Manjunatha S[*1], Meghashree A.C[*2]

[*1]PG Scholar, Department Of Digital Electronics And Communication Engineering

Visvesvaraya Technological University, Center For Post Graduate Studies, Mysuru, India.

[*2]Assistant Professor, Department Of Digital Electronics And Communication Engineering

Visvesvaraya Technological University, Center For Post Graduate Studies, Mysuru, India.

ABSTRACT

Using big data for prediction analysis along with machine learning or deep learning techniques or algorithms is one the most active areas of research in order to improve the health and the medical science. There is a significant increase in the size of the medical data as well as the complexity in the diagnosis of various diseases. With this being said, the diagnosis or the prediction of many terminal or fatal diseases has seen huge success through deep learning. Among those fatal diseases, pneumonia is one of the greatest threats to the life of a man affecting the lungs leading to lung failure. To diagnose a man with pneumonia, the x-ray of chest is needed, and an expert in the prediction is also required. Hence, it is more convenient to build an automated predictor to predict the pneumonia using the big data deep learning methods. Among all the other techniques, CNN (Convolutional Neural Networks) stand tall and high in this prediction along with other classifiers. Also, pre-training the CNN models for very large datasets that is for big data of healthcare units stands a high chance for accurate classification. A CNN model which is pre-trained along with three models Vgg, Resnet, Efficientnet an efficient feature extraction technique and various classifiers to classify the positive from negative is considered to give highly accurate results.

Keywords: Big Data, Deep Learning, Machine Learning, Pneumonia Prediction, CNN, Pre-Processing Of Dataset.

I.   INTRODUCTION

Pneumonia is caused by a bacterium and is a very dangerous disease which infects the lungs in turn affects the normal respiration of the human. This infection is a life-threatening one but specifically to infants and aged people over 60. In India alone, the cases recorded with pneumonia are 10 million per year and this tells the whole story about this deadly disease. Hence, the early prediction of the disease is much of use which leads us to finding suitable models to predict the disease. Having an expert to predict the pneumonia by observation of the x-ray is a bit of difficult as well as not to efficient method and also time consuming. So, for this prediction, making use of big data and deep learning is more effective as both deep learning and big data are rapidly growing fields. Big data refers to the complicated large datasets that can have an impact on the storage, security, manageability, manipulation and other renowned activities of traditional data warehousing. The big data was given a formal definition as ―Big Data is the information asset characterized by such a high Volume, Velocity and Variety to acquire specific technology and analytical methods for its transformation into Value‖. These are the important V's of big data. So in this prediction model we commonly use CNN to predict the X-ray sheets along with CNN we use models like Vgg, Resnet, Efficient net to find the accuracy.

II.   METHODOLOGY

This section provides detailed information on methodology of working and process flow of the proposed system which includes the basic operations like capturing data, processing of the capture data.

A diverse variety of photos is considered for the data set. These are normal x-ray photos of patients acquired by medical data. There will be both pneumonia-infected and normal pictures among them. The query image, which will be supplied as input, will be categorized based on these photographs from the database. Each picture will be labelled with either Pneumonia(P) or Normal (N). The size of the data set photos can be up to 1024*1024 pixels in this case, however the size of the input query image is limited to around 224*224 pixels in order to reduce computing complexity and gain faster computation.
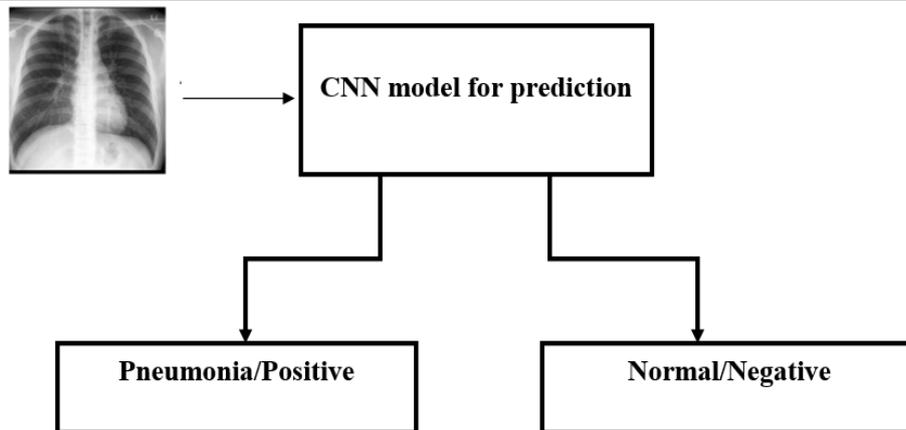
Fig 1: Prediction Model

While pre-training this same CNN model, feature extraction will be executed on the pictures in the considered dataset. The feature extraction process must be repeated for the input picture, which will be categorized eventually using multiple machine learning classifiers.

III.   SOFTWARE

A. Anaconda: Anaconda is a free and open-source machine learning application development tool that is often used. Anaconda is compatible with python and R, and may be used on Windows, Linux, and Mac OS X. We may install a python library using conda, which we refer to as the "anaconda package manager" in this programme. A company created by Peterwang and Oliphant, Anaconda, is responsible for its development and maintenance. Enterprise software is available from Anaconda, but it costs money.
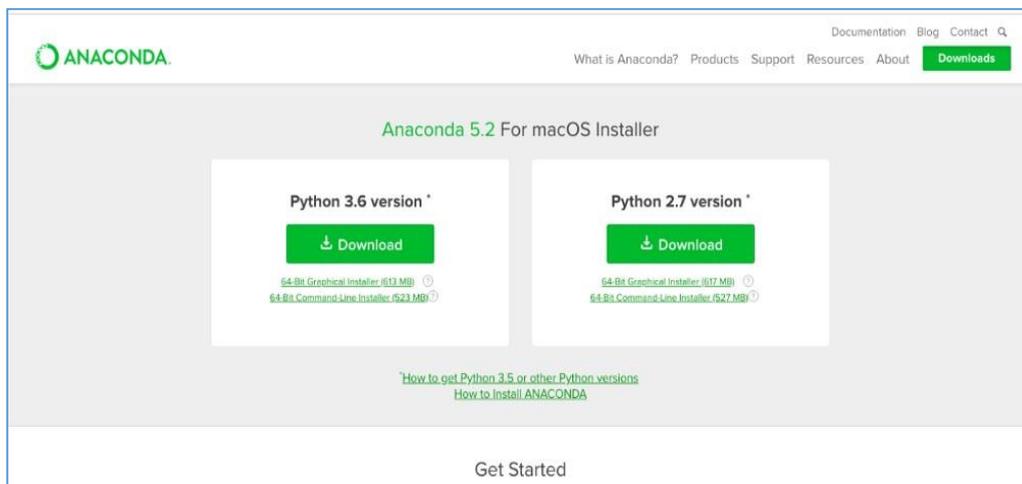
Fig 2: Anaconda Framework

To get started with Miniconda, you just need to install conda, and it doesn't include pip or any other package managers; it also has a limited amount of packages. Numpy, pandas, scikit-learn, and the jupyter notebook are some of the python packages included with the Conda environment manager.

B. PYTHON: Python, which is widely recognized as one of the most powerful programming languages for this particular task, is often used in the process of developing software for artificial intelligence. It is a basic high-level oops language that can also be used for general purpose programming. [C] is an acronym for Computer Operating System. Every single one of the Python programming courses, covering topics such as the "print" and "input" commands, strings, tuples, dictionaries, modules, file operations, data, and time.

C. Flask: Web Development Framework: The second most popular framework for developing micro-web applications is called Flask. Flask is used wherever there is a need to construct REST APIs. Other apps are then able to communicate with these application programming interfaces (APIs). The Flask Tutorial covers topics ranging from fundamentals to more advanced aspects of the Python Flask framework. Flask is a programming language that is simple to pick up and can be used by programmers of all experience levels.

D. Numpy: Numpy is the name of one of the most widely used libraries that can be found in Python. It is a large collection of high-level mathematical functions, One of the most typical applications of this library's capability is the creation of arrays. In addition to that, it is made up of items that are arranged in an array that has several dimensions.

E. Tensor Flow: The artificial intelligence discipline powerful machine learning is focused with the investigation of unstructured and semi - structured data. At its core, deep learning requires the processing of unstructured data in its raw form. Deep learning is an alternative to more traditional methods of machine learning that makes use of knowledge that has already been organized. Users of Tensor Flow have access to a vast and diverse set of libraries, tools, and resources provided by the community. It enables developers to build and deploy applications that are powered by cutting-edge machine learning technology, and it offers developers the capacity to do so.

IV. SYSTEM DESIGN

In order to get from a particular issue to a solution, the first step in the process is to design. Manager to begin the process of moving from the issue domain to the solution management, the problem must be defined. As a link between the development of requirements and the finished response, layout plays an important role here. The design method's goal is to provide a model or description of a system that may be used in the construction approach for that system. Known as a "gadget layout," this is the most recent variant. Systemic problem solving is one way to put this approach to work. The layout of a gadget is the most creative and challenging part of the whole process of making a device.
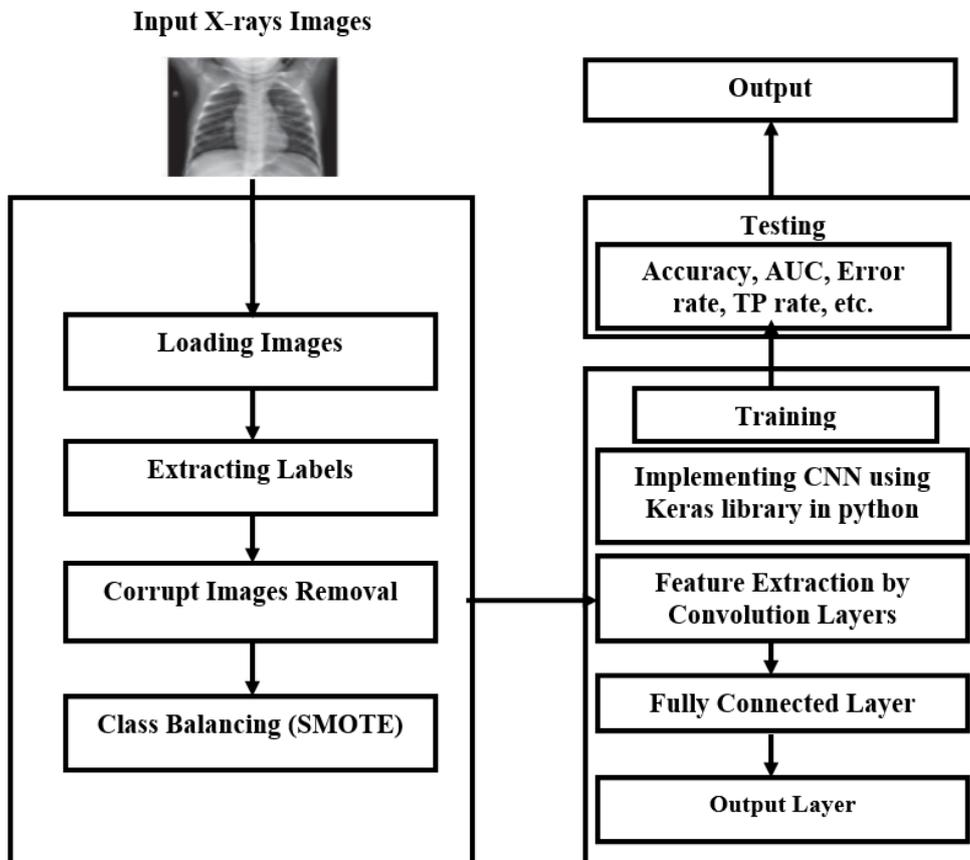

Fig 3: CNN Model Framework For Prediction Of Pneumonia

V. RESULTS

To Run the project in the command prompt with the python command python app.py.It run's the trained model which we had previously trained in the CNN model algorithm. That redirect to the local host URL https://127.0.0.1:5000.

Fig 4: URL of web page

When we click on the local host URL https://127.0.0.1:5000. It redirects to the main page of the Pneumonia project in which we can upload the sample X-ray image of the Pneumonia and predict the outcome of the positivity of Pneumonia.
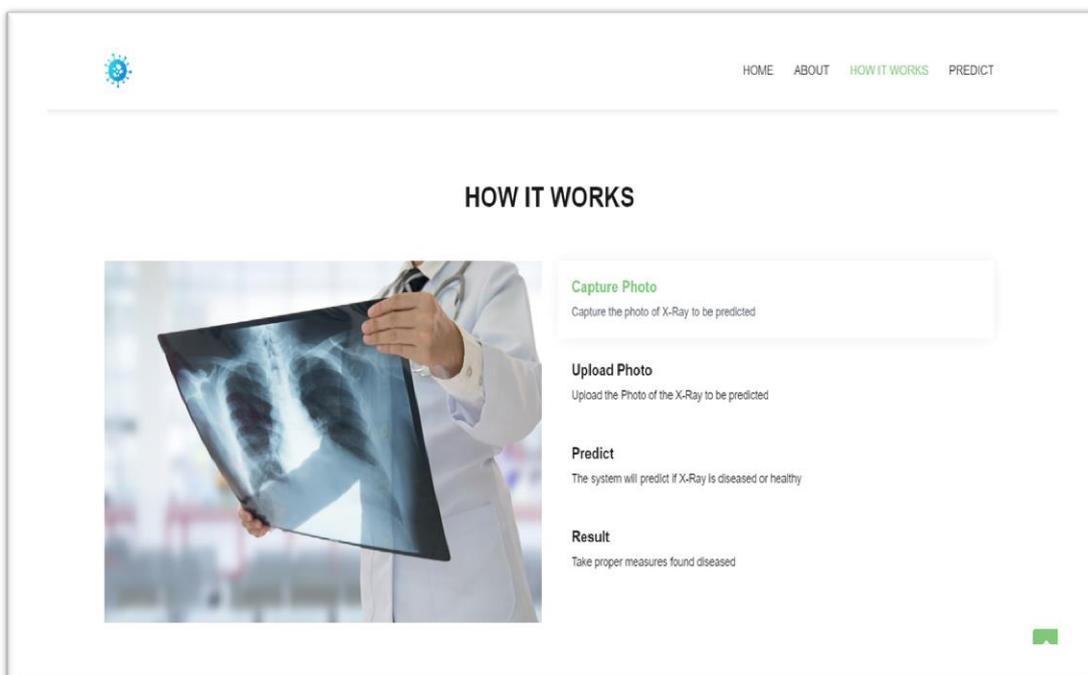

Figure 5: Pneumonia Project Main Page

In this Page we will upload the X-ray of the Pneumonia that will predict and classify the pneumonia disease of the chest X-ray. On selecting the image it will preedict the negetavie and postivity of the Pneumonia diseases of the X-ray.
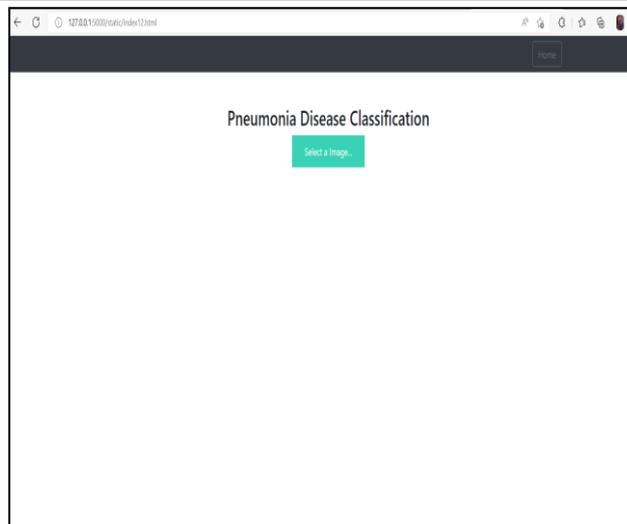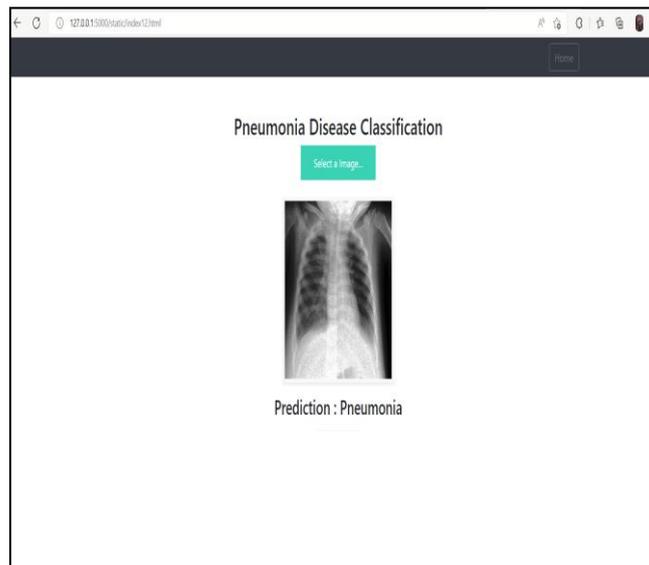
Figure 6: Pneumonia Disease Classification

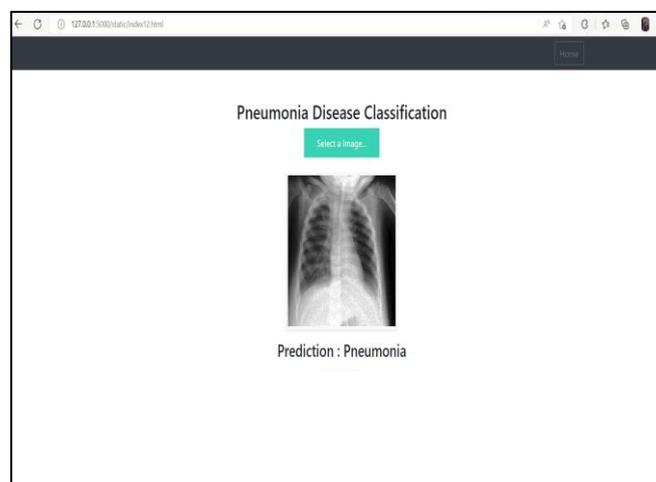**Figure 7:** Pneumonia Disease Classification of Negative X-Ray Sample

Figure 8: Pneumonia Disease Classification of Positive X-Ray Sample

In this project we had used the machine algorithm like VGG, resnet and efficientnet model. And by training and comparing all the three model, we got resnet Train model with the highest accuracy level
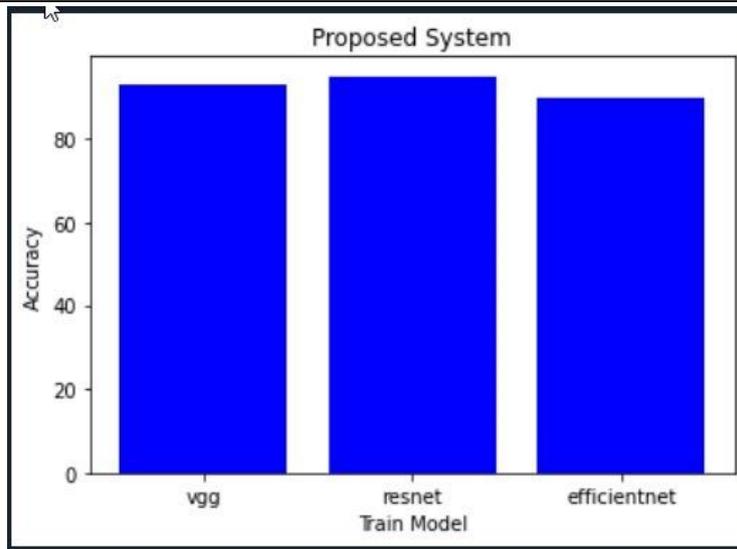
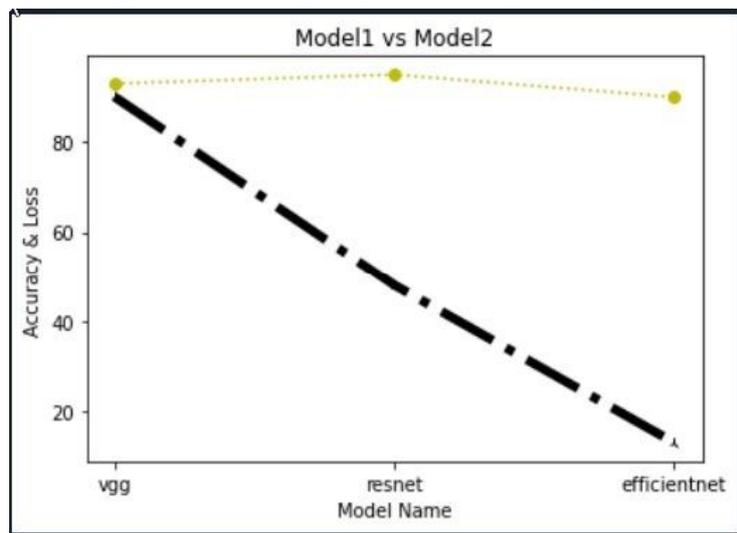**Figure 9:** Accuracy Graph of the Proposed System

Figure 10: Accuracy and loss graph of the trained models

This Graph represent the loss and accuracy level of the all three trained machine learning algorithm of VGG, resnet and efficientnet algorithms. We got resnet model with the highest level accuracy and with loss of the trained model.

VI. CONCLUSION

The CNN models are more accurate when compared any other model and the performance, efficiency is also high with the use of deep convolutional neural networks. While conducting repeated experiments with a greater number of different x-ray images, the performance was gradually increasing. The data re-sizing, as expected, had a quite good impact on the performance and the speed of computation. With varied sizes of images, varied accuracies and losses were obtained.

VII. REFERENCES

[1]   Ahmed, Z.; Mohamed, K.; Zeeshan, S.; Dong, X. (2020): Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, pp 1–35.

[2]   Chae, S.; Kwon, S.; Lee, D. (2018): Predicting Infectious Disease Using Deep Learning and Big Data. International Journal of Environmental Research and Public Health. 5(8): 1596, pp 1-20.

[3]   Crosta P. (2017): What you should know about pneumonia. Medical News Today.
        https://www.medicalnewstoday.com/articles/151632.

[4]   Gandomi A.; Murtaza H. (2015): Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), pp.137-144.

[5]   Goodfellow, I.; Bengio, Y.; Courville, A. (2016): Deep learning, in Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, p. 775.

[6]   Gupta, D.S. (2017): Architecture of Convolutional Neural Networks (CNNs) demystified. Analytics Vidya.

[7]   Jakhar, K.; Hooda, N. (2018): Big Data Deep Learning Framework using Keras: A Case Study of Pneumonia Prediction. 4th International Conference on Computing Communication and Automation (ICCCA).

[8]   Khan, N.; Yaqoob, I.; Hashem, I.A.T.; Inayat, Z.; Ali, W.K.M.; Alam, M., Shiraz, M.; Gani, A.; (2014): Big Data: Survey, Technologies, Opportunities, and Challenges. The Scientific World Journal, Article ID 712826, pp 1-18.

[9]   Mauro, A.D.; Greco, M.J.; Grimaldi, M. (2016): A formal definition of big data based on its essential features. Library Rev., 5 (3), pp. 122-135.

[10]   Najafabadi, M.M.; Villanustre, F.; Khoshgoftaar, T.M.; Seliya, N.; Wald, R.; Muharemagic E. (2015): Deep learning applications and challenges in big data analytics. Journal of Big Data, 2:1, pp 1-21.

[11]   Swetha K.R., Niranjanamurthy M., Amulya M.P., Manu Y.M.; (2021) Prediction of pneumonia using big data, deep learning and machine learning Techniques.

[12]   IEEE Xplore Part Number: CFP21AWO-ART.